Bayesian computation with Plug & Play priors for inverse problems in imaging sciences.

Rémi Laumont

Joint work with:
Valentin De Bortoli (ENS), Andrés Almansa, Julie Delon (Université Paris-Cité), Alain Durmus (Ecole Polytechnique) and Marcelo Pereyra (Heriot-Watt University).

AIP: September 5th, 2023

Introduction to the image restoration context.

True scene x Imaging device Observed image y

Figure: Forward problem

Model:

$$\boxed{y = A(x) + n}$$

with $x \in \mathbb{R}^d$ the unknown scene, $y \in \mathbb{C}^m$ the observation, $n \in \mathbb{C}^m$ the noise, and $A : \mathbb{R}^d \to \mathbb{C}^m$ a known degradation operator.

Introduction to the image restoration context.

True scene x Imaging device   Observed image y

Figure: Forward problem

Model:

$$\boxed{y = A(x) + n}$$

with $x \in \mathbb{R}^d$ the unknown scene, $y \in \mathbb{C}^m$ the observation, $n \in \mathbb{C}^m$ the noise, and $A : \mathbb{R}^d \to \mathbb{C}^m$ a known degradation operator.
Goal: Estimating x from its observation y. $\to$ ill-posed, ill-conditionned.

Bayesian paradigm.

- x is the realization of a random variable (r.v.) X on $\mathbb{R}^d$.
- y is the realization of a r.v. $Y|X = x$.
- Inferences about x from y are derived from the joint distribution (X, Y) via $p(x, y) = p(y|x)p(x)$.
- Posterior distribution computation via the Bayes's rule

$$\boxed{p(x|y) = \frac{p(x)p(y|x)}{\int_{\mathbb{R}^d} p(\tilde{x})p(y|\tilde{x})\mathrm{d}\tilde{x}} \propto p(x)p(y|x)}$$

where $p(x)$ is the prior and $p(y|x)$ is the likelihood (assumed to be known).

Classical estimators in imaging.

- Potential formulation: $R(x) = -\log p(x)$ and $F(x,y) = -\log p(y|x) (= \frac{\|Ax-y\|_2^2}{2\sigma^2})$
- Maximum-A-Posteriori (MAP) estimator:

$$\boxed{\hat{x}_{MAP} = \underset{x\in\mathbb{R}^d}{\arg\max}\; p(x|y) = \underset{x\in\mathbb{R}^d}{\arg\min} \left\{F(x,y) + \lambda R(x)\right\}.} \quad (1)$$

- Minimum Mean Square Error (MMSE) estimator:

$$\boxed{\hat{x}_{MMSE} = \underset{u\in\mathbb{R}^d}{\arg\min}\; \mathbb{E}[\|X-u\|^2 | Y=y] = \mathbb{E}[X|Y=y].} \quad (2)$$

Sampling using the Unadjusted Langevin Algorithm (ULA).

Goal: sampling from a distribution with target density $\pi(x) = p(x|y) \propto \exp(-R(x) - F(x,y))$.

Sampling using the Unadjusted Langevin Algorithm (ULA).

Goal: sampling from a distribution with target density $\pi(x) = p(x|y) \propto \exp(-R(x) - F(x,y))$.

- Langevin stochastical differential equation (SDE):

$$dX_t = \nabla \log \pi(X_t) dt + \sqrt{2} dB_t$$

with $(B_t)_{t \geq 0}$ a d-dimensional Brownian motion.

Sampling using the Unadjusted Langevin Algorithm (ULA).

Goal: sampling from a distribution with target density $\pi(x) = p(x|y) \propto \exp(-R(x) - F(x,y))$.

- Langevin stochastical differential equation (SDE):

$$dX_t = \nabla \log \pi(X_t)dt + \sqrt{2}dB_t$$

  with $(B_t)_{t\geq 0}$ a d-dimensional Brownian motion.

- ULA:

$$X_{k+1} = X_k + \delta\nabla \log \pi(X_k) + \sqrt{2\delta}Z_{k+1}$$

$$\boxed{X_{k+1} = X_k - \delta\nabla R(X_k) - \delta\nabla F(X_k, y) + \sqrt{2\delta}Z_{k+1}} \quad (3)$$

  with $Z_k \sim \mathcal{N}(0, Id)$ for all $k \in \mathbb{N}$ and $\delta > 0$.

Sampling using the Unadjusted Langevin Algorithm (ULA).

Goal: sampling from a distribution with target density $\pi(x) = p(x|y) \propto \exp(-R(x) - F(x,y))$.

- Langevin stochastical differential equation (SDE):

$$dX_t = \nabla \log \pi(X_t)dt + \sqrt{2}dB_t$$

  with $(B_t)_{t \geq 0}$ a d-dimensional Brownian motion.

- ULA:

$$X_{k+1} = X_k + \delta \nabla \log \pi(X_k) + \sqrt{2\delta} Z_{k+1}$$

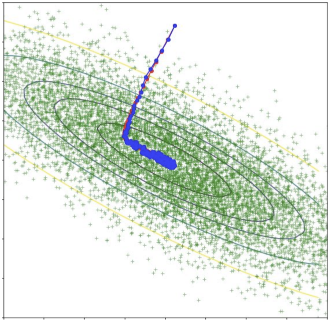
$$\boxed{X_{k+1} = X_k - \delta \nabla R(X_k) - \delta \nabla F(X_k, y) + \sqrt{2\delta} Z_{k+1}} \qquad (3)$$

  with $Z_k \sim \mathcal{N}(0, Id)$ for all $k \in \mathbb{N}$ and $\delta > 0$.

- Results (Durmus and Moulines, 2017):
  - Convergence towards a unique stationary distribution $\pi_\delta \approx \pi$ if $\nabla(R+F)$ is L-Lipschitz and $\delta < 1/L$.
  - Exponentially fast convergence if $F + R$ is strongly convex at ∞.

Sampling and Optimizing.

$$X_{k+1} = X_k - \delta \nabla R(X_k) - \delta F(X_k, y) + \begin{cases} \sqrt{2\delta} Z_{k+1} & \text{ULA.} \\ \delta Z_{k+1} & \text{SGD.} \end{cases}$$



Plug-and-Play (PnP) approaches.

Problem: $p(x)$ (or $R(x)$) is unknown and difficult to model.

Plug-and-Play (PnP) approaches.

Problem: $p(x)$ (or $R(x)$) is unknown and difficult to model.

- PnP methods use a denoiser $D_\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$ to implicitly define an image prior $p(x)$.
- Target prox_R (Hurault et al., 2022; Ryu et al., 2019; Venkatakrishnan et al., 2013).
- Target ∇R (Alain and Bengio, 2014), (Guo et al., 2019),(Kadkhodaie and Simoncelli, 2020) using Tweedie's formula (Miyasawa et al., 1961; Robbins, 1956).

Plug-and-Play (PnP) approaches.

Problem: $p(x)$ (or $R(x)$) is unknown and difficult to model.

- PnP methods use a denoiser $D_\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$ to implicitly define an image prior $p(x)$.
- Target prox_R (Hurault et al., 2022; Ryu et al., 2019; Venkatakrishnan et al., 2013).
- Target ∇R (Alain and Bengio, 2014), (Guo et al., 2019),(Kadkhodaie and Simoncelli, 2020) using Tweedie's formula (Miyasawa et al., 1961; Robbins, 1956).

In the literature:

- MAP point estimation: convergence towards fixed-points of some operators and/or under unrealistic assumptions.
- Sampling: No convergence guarantees.

Plug-and-Play (PnP) approaches.

Problem: $p(x)$ (or $R(x)$) is unknown and difficult to model.

- PnP methods use a denoiser $D_\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$ to implicitly define an image prior $p(x)$.
- Target prox_R (Hurault et al., 2022; Ryu et al., 2019; Venkatakrishnan et al., 2013).
- Target ∇R (Alain and Bengio, 2014), (Guo et al., 2019),(Kadkhodaie and Simoncelli, 2020) using Tweedie's formula (Miyasawa et al., 1961; Robbins, 1956).

In the literature:

- MAP point estimation: convergence towards fixed-points of some operators and/or under unrealistic assumptions.
- Sampling: No convergence guarantees.

Goal: Propose methods with convergence guarantees under realistic assumptions.

PnP approaches using Tweedies's formula.

Tweedie's formula

If $X \sim P_X$, $N \sim \mathcal{N}(0, Id)$ and $\tilde{X} = X + \sqrt{\varepsilon} N$ then,

$$\boxed{\mathbb{E}[X|\tilde{X} = \tilde{x}] - \tilde{x} = \varepsilon \nabla \log(p * g_\varepsilon)(\tilde{x}) = \varepsilon \nabla \log(p_\varepsilon)(\tilde{x})},$$

with g_ε a Gaussian kernel with variance ε.

PnP approaches using Tweedies's formula.

Tweedie's formula

If $X \sim P_X$, $N \sim \mathcal{N}(0, Id)$ and $\tilde{X} = X + \sqrt{\varepsilon} N$ then,

$$\boxed{\mathbb{E}[X|\tilde{X} = \tilde{x}] - \tilde{x} = \varepsilon \nabla \log(p * g_\varepsilon)(\tilde{x}) = \varepsilon \nabla \log(p_\varepsilon)(\tilde{x})},$$

with g_ε a Gaussian kernel with variance ε.

- $p(x)$ is unknown but $p_\varepsilon(x)$ can be used.

PnP approaches using Tweedies's formula.

Tweedie's formula

If $X \sim P_X$, $N \sim \mathcal{N}(0, Id)$ and $\tilde{X} = X + \sqrt{\varepsilon}N$ then,

$$\boxed{\mathbb{E}[X|\tilde{X} = \tilde{x}] - \tilde{x} = \varepsilon \nabla \log(p * g_\varepsilon)(\tilde{x}) = \varepsilon \nabla \log(p_\varepsilon)(\tilde{x})},$$

with g_ε a Gaussian kernel with variance ε.

- $p(x)$ is unknown <u>but</u> $p_\varepsilon(x)$ can be used.
- Using the MMSE denoiser $D^*_\varepsilon(\tilde{x}) = \mathbb{E}[X|\tilde{X} = \tilde{x}]$ we get

$$\boxed{\nabla \log(p_\varepsilon)(\tilde{x}) = (D^*_\varepsilon(\tilde{x}) - \tilde{x})/\varepsilon} \ ,$$

  with $\nabla R \approx -\nabla \log p_\varepsilon$.
- <u>Problem:</u> $D^*_\varepsilon = ?$

PnP approaches and Tweedie's formula.

- PnP-SGD (Laumont et al., 2022b):

$$X_{k+1} = X_k - \delta_k \nabla F(X_k, y) + \delta_k (D_\varepsilon(X_k) - X_k)/\varepsilon) + \delta_k Z_{k+1},$$

where $(\delta_k)_{k\in\mathbb{N}}$ is a sequence of decreasing step-sizes.
$\rightarrow$ it converges in the vicinity of the stationary points of $\log p(x|y)$.

PnP approaches and Tweedie's formula.

- PnP-SGD (Laumont et al., 2022b):

$$X_{k+1} = X_k - \delta_k \nabla F(X_k, y) + \delta_k (D_\varepsilon(X_k) - X_k)/\varepsilon) + \delta_k Z_{k+1},$$

where $(\delta_k)_{k\in\mathbb{N}}$ is a sequence of decreasing step-sizes.
$\rightarrow$ it converges in the vicinity of the stationary points of $\log p(x|y)$.

- PnP-ULA (to sample from $\pi^C_{\delta,\varepsilon}$) (Laumont et al., 2022a):

$$X_{k+1} = X_k - \delta \nabla F(X_k, y) + \delta (D_\varepsilon(X_k) - X_k)/\varepsilon) + \delta (\Pi_C(X_k) - X_k)/\lambda + \sqrt{2\delta} Z_{k+1}.$$

where this term ensures the strong convexity in the tails and Π_C is a projection on $B(0, R_C)$ and $D_\varepsilon(x) \simeq D^*_\varepsilon(x)$.

Plug-and-Play approaches and Tweedie's formula.

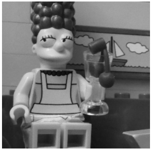
- Hypotheses:
  - D_ε is Lipschitz and there exists $M : \mathbb{R}^+ \to \mathbb{R}^+$, such that for all $\|x\| \leq R$, $\|D_\varepsilon(x) - D_\varepsilon^*(x)\| \leq M(R)$.
  - The likelihood $p(y|x)$ is bounded, C^1 and $\nabla \log p(y|x)$ is Lipschitz.
  - The MSE loss for D_ε^* under $g_\varepsilon(.|\tilde{x})$ is finite and uniformly bounded.
- Non-asymptotic error when sampling (Laumont et al., 2022a):

$$|\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_{\pi_{\delta,\varepsilon}^C}[X_k] - \int_{\mathbb{R}^d} \tilde{x} p(\tilde{x}|y) \mathrm{d}\tilde{x}| \leq C_0\{C_1 \varepsilon^{\beta/4} + C_2 R_C^{-1} + C_3(\sqrt{\delta} + \frac{1}{n\delta} + C_R)\}. \quad (4)$$

Problem position: $y = Ax + n$

- **Deblurring:** A encodes a block filter of size 9 and $\sigma = 1/255$.
- **Interpolation:** A is a diagonal matrix with 1 or 0 on the diagonal and hiding 80% of the pixels in the original image.
- **Dataset:**

Cameraman.



Simpson.

Traffic.

Alley.

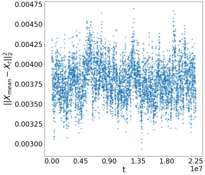
Bridge.

Goldhill.

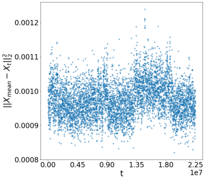
Algorithm parameters for PnP-ULA and PnP-SGD

- **Denoising implicit prior D_ε:** SN-DnCNN provided by (Ryu et al., 2019) and such that $(D_\varepsilon - Id)$ is L-Lipschitz with $L < 1$. $\varepsilon = (5/255)^2$.
- **PnP-ULA:**
  - Initialization at the observation y.
  - Number of iterations $n = 2.5e7$.
  - $\delta = \delta_{\texttt{stable}}$
  - $C = [-1, 2]^d$.
- **PnP-SGD:**
  - Initialization with TV-L2.
  - Number of iterations $n = 5e3$ after the burn-in.
  - $\delta_0 = \delta_{\texttt{stable}}/6$ and $\delta_k = \delta_0(k + 1 - n_{\texttt{burn-in}})^{-0.8}$.

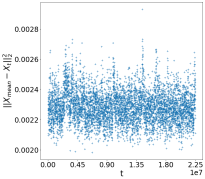
Evolution of the L_2 distance between the final MMSE estimate and the samples generated by PnP-ULA for the interpolation problem.



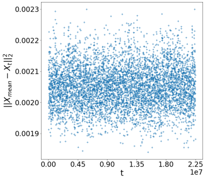
Cameraman.



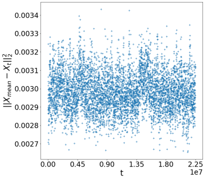
Simpson.



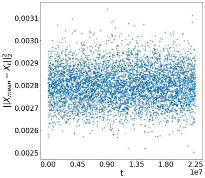
Traffic.



Alley.

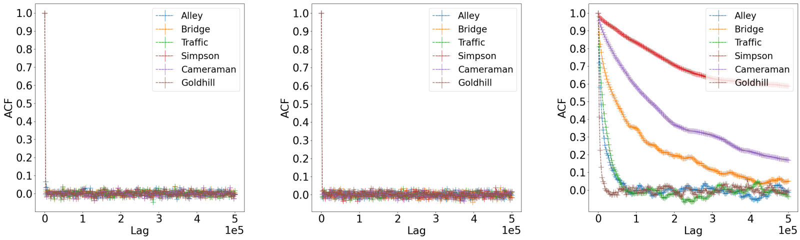


Bridge.



Goldhill.

ACF in the pixel domain for the interpolation problem.

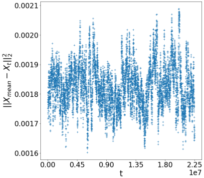


Fastest direction

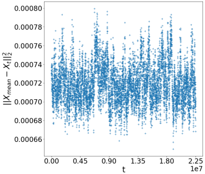
Median direction

Slowest direction

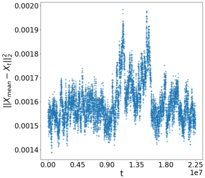
Evolution of the L_2 distance between the final MMSE estimate and the samples generated by PnP-ULA for the deblurring problem.



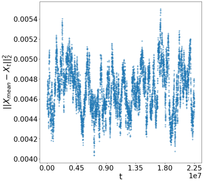
Cameraman.



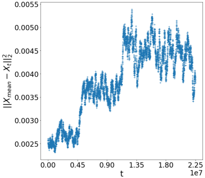
Simpson.



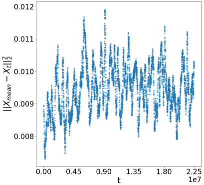
Traffic.



Alley.

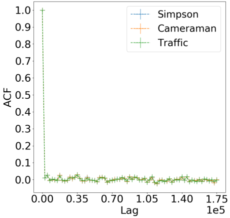


Bridge.

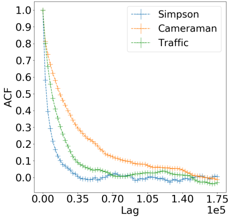


Goldhill.

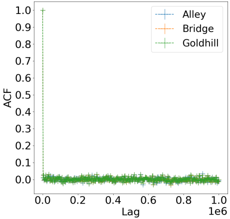
ACF in the Fourier domain for the deblurring problem.



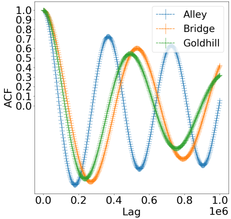
Fastest direction



Slowest direction

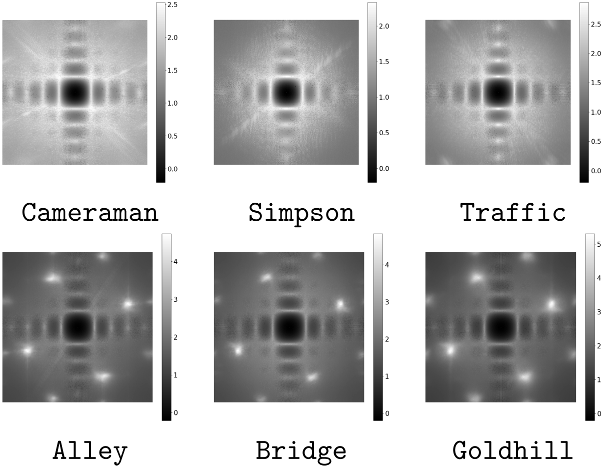


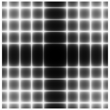
Fastest direction



Slowest direction

Log-standard deviation maps in the Fourier domain for the Markov chains defined by PnP-ULA for the deblurring problem.

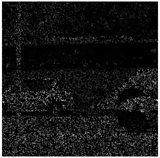




Inverse Fourier transform of the blur kernel.

Observation.





PSNR=6.69/SSIM=0.11 PSNR=7.43/SSIM=0.04 PSNR=8.35/SSIM=0.09

PnP-ULA.

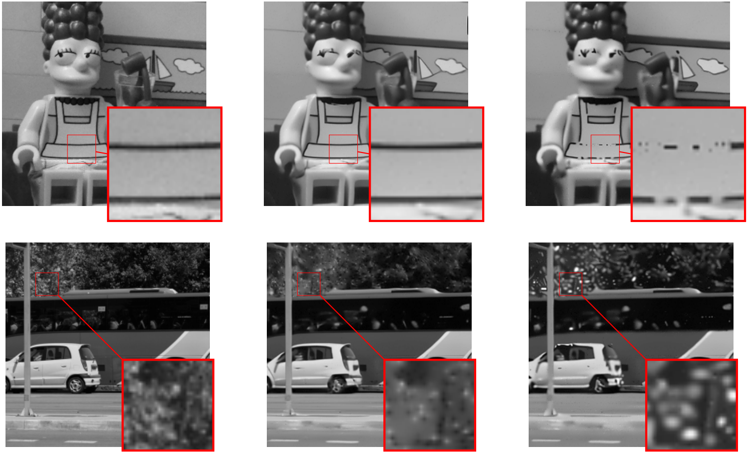


PSNR=25.06/SSIM=0.89 **PSNR=30.62/SSIM=0.93** **PSNR=26.90/SSIM=0.85**

PnP-SGD, init=TVL2.

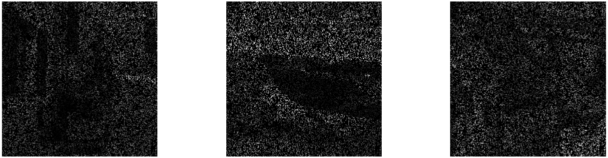


PSNR=23.94/SSIM=0.88 PSNR=28.90/SSIM=0.90 PSNR=24.20/SSIM=0.81



Original image. PnP-ULA. PnP-SGD.

Observation.



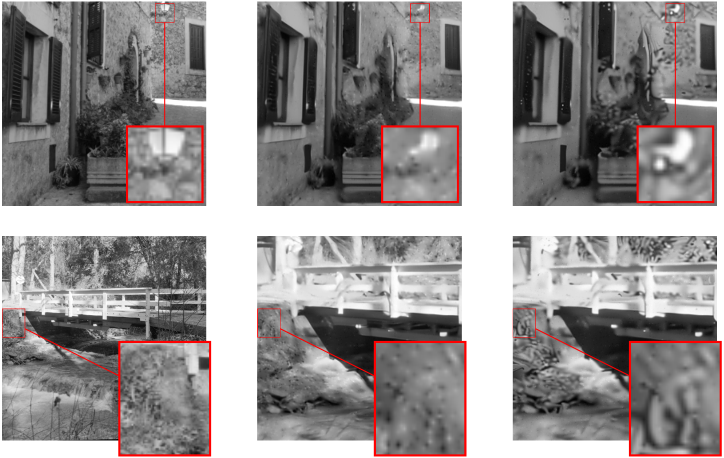
PSNR=8.27/SSIM=0.004 PSNR=5.71/SSIM=0.004 PSNR=6.61/SSIM=0.03

PnP-ULA.

PSNR=27.74/SSIM=0.79 **PSNR=26.16/SSIM=0.80** **PSNR=26.76/SSIM=0.74**

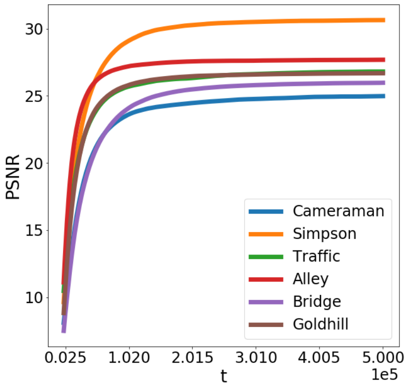
PnP-SGD, init=TVL2.

PSNR=26.45/SSIM=0.75 PSNR=24.71/SSIM=0.77 PSNR=25.96/SSIM=0.72

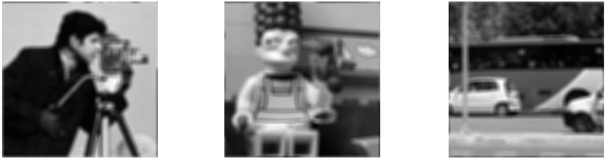


Original image. PnP-ULA. PnP-SGD.

PSNR evolution of the estimated MMSE for the interpolation problem with PnP-ULA.



Observation

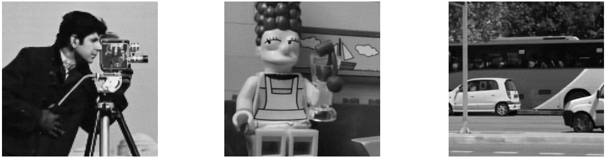


PSNR=20.30/SSIM=0.70 | PSNR=22.44/SSIM=0.66 | PSNR=20.34/SSIM=0.49

PnP-ULA.

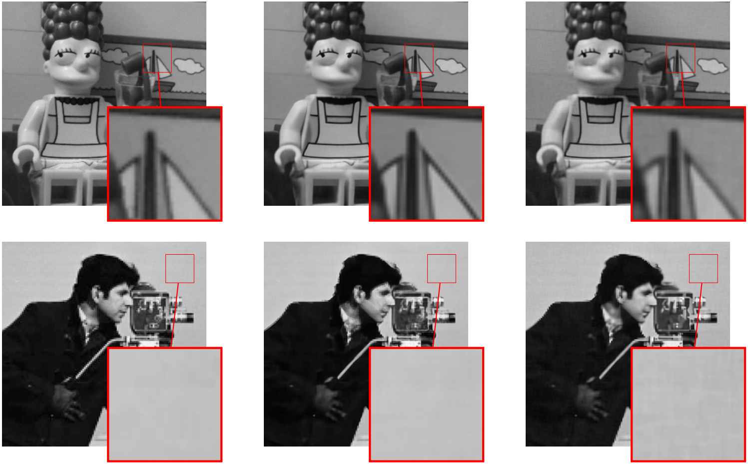
PSNR=30.50/**SSIM=0.93** | **PSNR=34.26/SSIM=0.94** | **PSNR=29.90/SSIM=0.90**

PnP-SGD.



PSNR=30.73/SSIM=0.92 | PSNR=33.52/SSIM=0.92 | PSNR=29.42/SSIM=0.88

Detailed comparison between PnP-ULA and PnP-SGD.



Original image. PnP-ULA. PnP-SGD.

Observation.

PSNR=22.64/SSIM=0.46

PSNR=21.84/SSIM=0.49

PSNR=22.61/SSIM=0.45

PnP-ULA.

PSNR=28.98/SSIM=0.80

PSNR=28.28/SSIM=0.84

PSNR=27.72/SSIM=0.73

PnP-SGD.

PSNR=29.26/SSIM=0.82

PSNR=28.04/**SSIM=0.84**

PSNR=28.27/SSIM=0.76

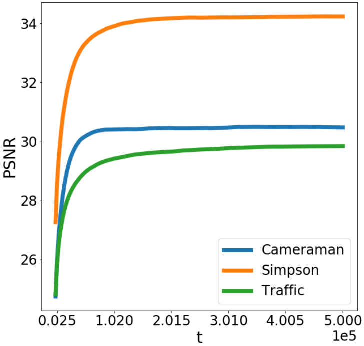
Detailed comparison between PnP-ULA and PnP-SGD.

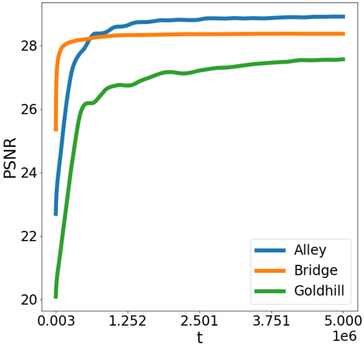
Original image.

PnP-ULA.

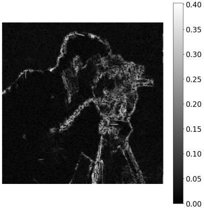
PnP-SGD.

PSNR evolution of the estimated MMSE for the deblurring problem with PnP-ULA.

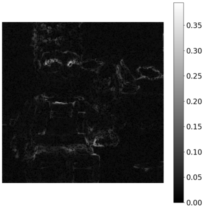




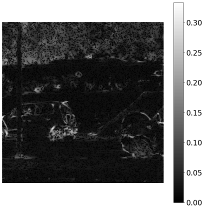
Marginal posterior standard deviation of the unobserved pixels for the interpolation problem.



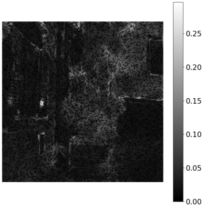
Cameraman



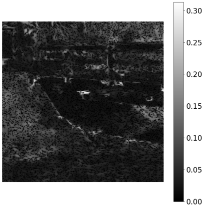
Simpson



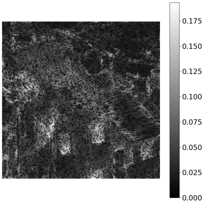
Traffic



Alley

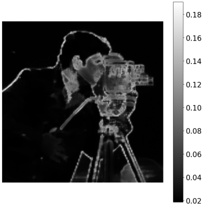


Bridge

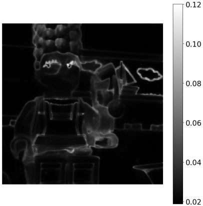


Goldhill

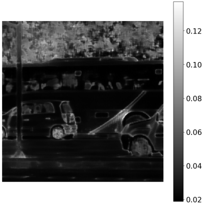
Marginal posterior standard deviation for the deblurring problem.



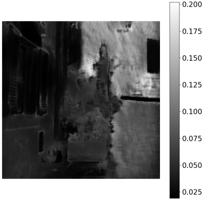
Cameraman



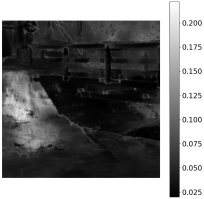
Simpson



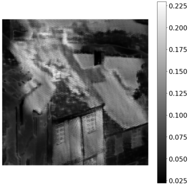
Traffic



Alley

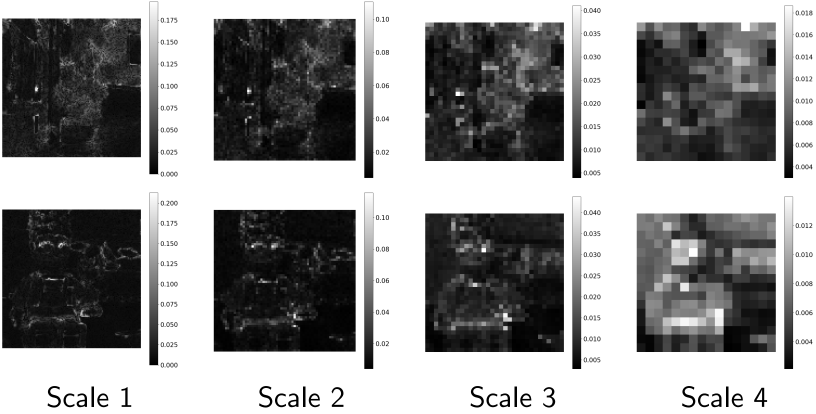


Bridge



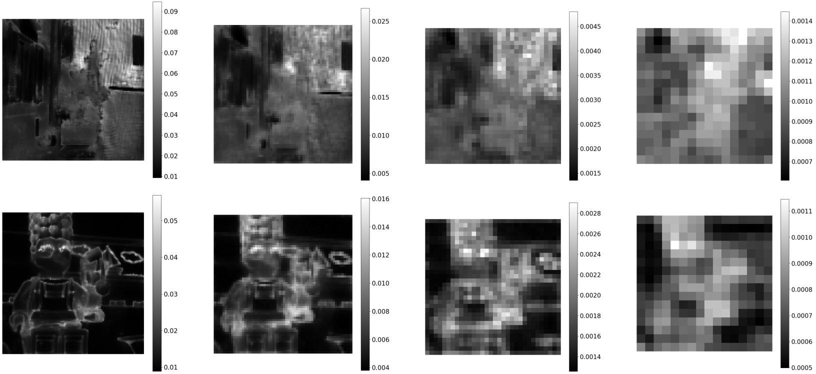
Goldhill

Standard deviation for `Alley` and `Simpson` images for interpolation at different scales.



The scale i corresponds to a downsampling by a factor $2i$ of the original sample size.

Standard deviation for `Alley` and `Simpson` for deblurring at different scales.



Scale 1 | Scale 2 | Scale 3 | Scale 4

The scale i corresponds to a downsampling by a factor $2i$ of the original sample size.

Conclusion.

- Summary.
  - Better understanding of the PnP models for point estimation and sampling.
  - Development of efficient Langevin based algorithms with detailed convergence guarantees under realistic hypothesis.
- Future work.
  - Development of accelerated schemes for sampling from the posterior distribution.
  - Correcting artefacts generated by the prior denoiser.
  - Study of more realistic and challenging inverse problems (CT reconstruction, semi-blind deblurring, etc, ...).

Publications.

- Laumont, R., De Bortoli, V., Almansa, A., Delon, J., Durmus, A., & Pereyra, M. (2022). Bayesian imaging using Plug & Play priors: when Langevin meets Tweedie. SIAM Journal on Imaging Sciences, 15(2), 701-737.
- Laumont, R., De Bortoli, V., Almansa, A., Delon, J., Durmus, A., & Pereyra, M. (2023). On Maximum a Posteriori Estimation with Plug & Play Priors and Stochastic Gradient Descent. Journal of Mathematical Imaging and Vision, 1-24.

THANK YOU.

Alain, Guillaume and Yoshua Bengio (2014). "What Regularized Auto-Encoders Learn from the Data-Generating Distribution". In: *Journal of Machine Learning Research* 15, pp. 3743–3773. ISSN: 1532-4435. arXiv: 1211.4246 (cit. on pp. 12–15).

Durmus, Alain and Éric Moulines (2017). "Nonasymptotic convergence analysis for the unadjusted Langevin algorithm". In: *Ann. Appl. Probab.* 27.3, pp. 1551–1587. ISSN: 1050-5164. DOI: 10.1214/16-AAP1238 (cit. on pp. 7–10).

Guo, Bichuan, Yuxing Han, and Jiangtao Wen (2019). "AGEM: Solving Linear Inverse Problems via Deep Priors and Sampling". In: *Advances in Neural Information Processing Systems*, pp. 547–558 (cit. on pp. 12–15).

Hurault, Samuel, Arthur Leclaire, and Nicolas Papadakis (2022). "Gradient Step Denoiser for convergent Plug-and-Play". In: *International Conference on Learning Representations* (cit. on pp. 12–15).

Kadkhodaie, Zahra and Eero P Simoncelli (2020). "Solving Linear Inverse Problems Using the Prior Implicit in a Denoiser". In: *arXiv preprint arXiv:2007.13640* (cit. on pp. 12–15).

Laumont, Rémi, Valentin De Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra (2022a). "Bayesian Imaging Using Plug & Play Priors: When Langevin Meets Tweedie". In: *SIAM Journal on Imaging Sciences* 15.2, pp. 701–737. DOI: 10.1137/21M1406349 (cit. on pp. 19–21).

Laumont, Rémi, Valentin de Bortoli, Andrés Almansa, Julie Delon, Alain Durmus, and Marcelo Pereyra (2022b). "On Maximum-a-Posteriori estimation with Plug & Play priors and stochastic gradient descent". In: (cit. on pp. 19, 20).

Miyasawa, Koichi et al. (1961). "An empirical Bayes estimator of the mean of a normal population". In: *Bull. Inst. Internat. Statist* 38.181-188, pp. 1–2 (cit. on pp. 12–15).

Robbins, Herbert (1956). "An Empirical Bayes Approach to Statistics". In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Vol. 3. University of California Press, pp. 157–164 (cit. on pp. 12–15).

Ryu, Ernest K., Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin (2019). "Plug-and-Play Methods Provably Converge with Properly Trained Denoisers". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, pp. 5546–5557. arXiv: 1905.05406 (cit. on pp. 12–15, 23).

Venkatakrishnan, Singanallur V, Charles A Bouman, and Brendt Wohlberg (2013). "Plug-and-play priors for model based reconstruction". In: *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, pp. 945–948 (cit. on pp. 12–15).